

心的負荷状況における車載情報機器のための音声対話戦略の分析

野田 幸志[†] 西田 昌史[‡] 堀内 靖雄[‡] 市川 薫[‡]

[†] [‡] 千葉大学大学院自然科学研究科 〒263-8522 千葉県稲毛区弥生町 1-33

E-mail: [†] noddy@graduate.chiba-u.jp [‡] {nishida,hory,ichikawa}@faculty.chiba-u.jp,

あらまし 我々は、対話による車載情報機器の音声インタフェースの構築を目指している。運転時においては、安全性、即時性といった観点からユーザに心的負荷をかけずにシステムとの効率的な音声対話を実現する必要がある。そこで、本研究では、カーナビゲーションシステムの飲食店検索に場面を設定し、システムが音声で提示する情報量やユーザの目的意識の違いなどに着目して、対話時における心的負荷の影響について検討を行った。音声対話はWoZ法を用いて行い、二重課題法とNASA-TLXにより対話時の負荷量について分析を行った。その結果、音声で提示する情報量が多く、ユーザの目的意識が低い場合に対話時の負荷量が大きくなることが明らかとなった。

キーワード 心的負荷, 車載情報機器, 音声対話戦略, 二重課題法, NASA-TLX,

Analysis of spoken dialogue strategy for in-vehicle information system under mental workload condition

Kouji NODA[†] Masafumi NISHIDA[‡] Yasuo HORIUCHI[‡] and Akira ICHIKAWA[‡]

[†] [‡] Graduate School of Science and Technology, Chiba University

1-33 Yayoi-cho, Inage-ku, Chiba, 263-8522 Japan

E-mail: [†] noddy@graduate.chiba-u.jp [‡] {nishida,hory,ichikawa}@faculty.chiba-u.jp,

Abstract We aim to construct a spoken dialogue system for in-vehicle information system. It is necessary to achieve an efficient spoken dialogue by controlling a mental workload from the viewpoint of safety and immediateness when a driver uses a speech interface under driving condition. In this study, we focused on user's sense of purpose and amount of output information from a system in a restaurant retrieval of a car navigation system and studied the mental workload of the spoken dialogue when a driver uses an in-vehicle information system by speech interface. We conducted spoken dialogue experiments based on the WoZ method and analyzed the mental workload by using dual-task method and NASA-TLX. As a result, it demonstrated that the mental workload increased when the amount of speech information by the system were large and user's sense of purpose were low.

Keyword Mental workload, In-vehicle information system, Spoken dialogue strategy, Dual-task, NASA-TLX

1. はじめに

音声認識・音声合成の技術の進展により、様々な音声対話システムが実用化され始めている。例えば、カーナビやカーオーディオなど車載情報機器への音声対話インタフェース実現の要求がある。これは、音声対話インタフェースが、指でのボタンやキー入力が必要としない（ハンズフリー）、ディスプレイなどに視覚的な注意を払わずに使用できる（アイズフリー）という利点を持ち、運転時に車載情報機器を操作することにより起こりうる脇見運転での事故のリスクを抑えるのに有効と考えられるためである。更に、車載情報機器において、運転のような心的負荷がかかっている状況を考慮したシステム応答をすることで、ユーザにとって使いやすいシステムとなると考えられる。このような、協調的な音声対話システムの実現について、ユーザの嗜好やシステムへの慣れの状況に適應させる研究

が盛んに進められている[1]。しかし、協調的な音声対話システムの対話戦略の評価について、対話ターン数や対話達成率、アンケートによる主観評価など利便性に関する評価が主であるが、運転状況を考慮する場合、利便性に加え、対話が運転中のユーザに与える影響を十分に検証する必要がある。

音声対話システムがユーザに与える影響を負荷（Workload）として考えた場合、対話などの作業に対する負荷量を測定する手法として、二重課題法やNASA-TLXが使われている。二重課題法などを用いて、システムとの対話が与える影響を負荷量として測定した従来研究として、宗近ら[2]は、対話の思考拘束形態の違いでの負荷量の差を調べるため、対話を連想・計算・記憶に分類し、それぞれ、しりとり対話・暗算対話・個人情報の質疑応答対話、などで負荷量の差を測定している。脇田ら[3]清水ら[4]は、交通情報検索の対

話において、対話形式の違い（単語列挙・定型文・Slot Filling）で負荷量を測定している。西本ら[5]は、飲食店検索対話において、『条件絞込み』や『候補選択』など各対話場面で負荷量の差を測定し、ユーザに大きな負荷を与える対話場面を特定できることを実証した。

これら従来研究より、一発話で多くの思考（覚える・思い出す・判断する・など）を要求するような、選択肢（提示される情報量）の多いシステムの発話が負荷の要因の一つと考えられる。つまり、システム側が情報の種類・量を絞って伝える対話戦略が心的負荷状況下では有効であると思われる。しかし、運転のような負荷状況ではない、平常時の音声対話での適切な提示量を満足度から検証した研究[6]や、音声で暗記する単語量が運転に与える影響を検証した研究[7]はあるものの、一発話で提示する情報量の違いが心的負荷状況のユーザに与える影響を検証した研究はない。また、一発話で提示される情報量を少なくすることは、対話ターン数を増加させる要因になり、利便性に悪い影響を与えると思われる。更に、利便性の悪さがユーザにフラストレーションを感じさせ、負荷量が対話全体を通して増加する可能性も考えられる。

そこで我々は、利便性に加え、ユーザに負担をかけない音声対話戦略の提案と、提案を実装した音声対話システムの実現を目指し、一発話で提示する情報量が、心的負荷状況のユーザに与える影響を明らかにする。実験に用いる対話システムでは WoZ 法を用い、認識率の影響を排除して評価する。対話タスクとしては、飲食店検索対話を用いる。検索対話では、ユーザの目的意識が高い場合（ある程度、具体的な候補が存在）と低い場合（大まかな検索条件のみ）があり、それぞれ対話の質が異なり、目的意識が低い場合の方がより負荷が高いと考えられる。そこで、提示する情報量に加え、目的意識の違いについても負荷量測定し、目的意識に応じた対話戦略の検討を行う。

本論文では、2章では実験に用いる負荷測定法について説明し、3章で状況（運転・目的意識）と対話戦略（音声対話システムが一発話で提示する情報量）がユーザに与える影響を検証する実験の説明・結果・考察について述べる。そして最後に、4章でまとめと今後の課題について述べる。

2. 負荷測定法

2.1. 二重課題法

二重課題法は、被験者に異なる2つの課題を同時に遂行させ、課題遂行能力を指標として、情報処理資源の容量や資源の容量配分の特性を評価するものであり、負荷量を客観値で測定することができる。

音声対話の負荷量測定を目的とする場合、対話課題

に加えて、負荷課題として、反応課題や運転課題などが用いられている。反応課題は、視覚刺激が発生してからキー入力やブレーキを踏むまでの反応時間や反応遅れ率を負荷量測定の指標にしている。脇田ら[3]、清水ら[4]、内山[7]、萩原ら[8]は実車やドライビングシミュレータを、西本ら[5]、篠原ら[9]、工藤ら[10]は簡易的なアプリケーションを用いて実験している。運転課題は、ドライビングシミュレータを用い、指示された運転位置からのずれ（走行誤差）を、負荷量を測定する指標としている。宗近ら[2]、清信ら[11]は走行誤差で負荷量を測定している。

本研究では、心的負荷状況を実験的に再現するため反応課題を用いる。反応課題の画面構成を図1に示す。画面には運転者の視界を模した動画を表示する。そして、指定したキー入力を促す指示キーが画面上の様々な場所に、不規則に表示される。被験者は指示キーの出現に反応して、指示キーと同じ向きのキーボード矢印キーを入力する。指示キーが表示されてから矢印キーの入力までの時間を反応時間とし、その値から反応遅れ率を求め、負荷量の指標とする。反応遅れ率とは、脇田ら[2]を参考に、対話課題と同時に実施する負荷課題での反応時間が、負荷課題単体の平均反応時間と標準偏差の和を超えた割合であり、反応時間での評価に比べ、個人差を相殺できると考えられる。



図1 反応実験の画面構成

2.2. NASA-TLX

NASA-TLXは、作業の心的負荷を主観値で評価する方法である[13]。NASA-TLXでは、課題の負荷を求めするために、6つの評価尺度を用いる。被験者には、この6項目について評価させた後、各評価尺度に対して重みをつけさせることで、課題に対する総合的な負荷を求める。6つの評価尺度は、知覚的要求、身体的要求、タイムプレッシャー、作業成績、努力、フラストレーションで、評価の際は被験者に対して、各評価尺度を文章で説明する。評価の手順は以下の通りである。

- (1) 評価するタスクにおいて、6つの評価尺度に対して、段階評価する。
- (2) 6つの評価尺度が作業の負荷に対する影響

が大きいかどうか一対比較により評価する。

(3) 一対比較により得られた重要度の重みに従い、評価尺度を総合した負荷の値を求める。

また、NASA-TLX では、6つの尺度の評価から心的負荷の詳細な要因を分析することが可能である。

萩原ら[9]は運転中における携帯電話のハンズフリーモードでの対話によって受ける心的負荷量の測定と、対話内容や年齢による負荷量の差を、中園ら[12]は手話対話者が手話画像の遅延によって受ける心的負荷量を、NASA-TLX を用いて測定している。特に、萩原らは二重課題法による反応時間と NASA-TLX 総合値の相関性を確認している。一方、宗近ら[2]も対話内容の違いによる負荷量の差を NASA-TLX を用いて測定しているが、走行誤差との相関は見られなかった。考えられる要因の一つとして、萩原らや中園らは、被験者が6項目の評価尺度を十分に理解できるように、日本語版 NASA-TLX を改善し、評価尺度の説明を実験に合わせて簡易化および具体化している点が挙げられる。

そこで本研究でも、従来研究を参考に、用いる6つの評価尺度の段階評価アンケートに用いる尺度の説明文と、段階評価の終端点の説明を、音声対話システムとの対話に合わせて具体的かつ簡潔にして、被験者に直感的に評価できるようにする(表1・図2)。

表1 NASA-TLX の評価尺度の説明文

尺度名	説明文	終端点(最小)の説明	終端点(最大)の説明
知覚的要求	考えたり、覚えたりするのは大変でしたか	単純だった。覚えきれた。	複雑だった。覚え切れなかった。
身体的要求	身体は疲れましたか。	動作がらくだった。休み休みできた。	動作がきつかった。動きっぱなしだった
タイムプレッシャー	あせりましたか。	余裕があった。ゆっくりできた。	余裕がなかった。速かった。
作業成績	上手くやり遂げられましたか。	満足している。達成できた。	満足していない。十分は達成できず。
努力	一生懸命になりましたか。	努力する程ではなかった。	かなり努力した。
フラストレーション	イライラしましたか。	リラックスしていた。	ストレスを感じた。

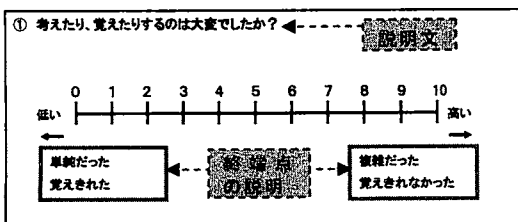


図2 尺度評価アンケート例(知覚的要求)

3. 実験

提示量と目的意識の異なる実験条件で、音声対話システムとの対話が被験者に与える負荷量を測定し、負荷の要因を分析するため、2章で紹介した二重課題法と NASA-TLX を用い、対話課題と同時に心的負荷状況として反応課題を被験者に実施させる。

3.1. 対話課題

対話課題として、音声対話システムと飲食店検索対話を被験者に行ってもらう。音声対話システムに関して、今回は WoZ 法を用いる。これにより、音声認識率の影響を排除した評価ができる。

被験者には「音声で提示される店名から行きたいお店を決めて、システムに伝える」という課題を与える。提示される情報量の違いを比較するため、システムに飲食店の検索条件を入力し終了時点からの対話に場面を限定し、一発話で提示する店名数を情報量の差とする。そして、平常時の音声対話で許容される提示数は5個程度、という従来研究[7]を参考に、心的負荷状況では許容できる提示数が下がると予想されるので、提示する店名数は、P1(1店ずつ)、P2(2~3店ずつ)、P3(4~7店ずつ)、の3パターンで実験する。

[検索条件] 和食, 安い [店舗候補] カップ寿司
 ++++++

(S: システム発話 U: ユーザ発話)

S1: 和食, 値段安いで検索しますね。
 U1: はい。
 S2: 候補が10件。大戸屋はどうですか?
 U2: 違うのは?
 S3: 花丸うどんはどうですか?
 U3: ほかは?
 S4: カップ寿司はどうですか?
 U4: そこにします。

図3 対話例1: 目的意識[高]・提示量1

[検索条件] ファーストフード, 近い
 [店舗候補] 自由
 ++++++

S1: ファーストフード, 距離近いで検索しますね。
 U1: はい。
 S2: 候補が10件。マクドナルド, モスバーガー ロッテリアなどはどうですか?
 U2: 他には?
 S3: フレッシュネスバーガー, サブウェイ, ファーストキッチンなどはどうですか?
 U3: じゃあ, モスバーガーでお願い。

図4 対話例2: 目的意識[低]・提示量2~3

また、1章で述べたように検索対話では、ユーザの目的意識が高い場合と低い場合があり、それぞれ対話の質が異なると考えられる。車内での飲食店検索対話で例を挙げると、ある程度、行きたいお店の候補が頭に浮かんでいる状況での対話が、目的意識が高い対話で、おなかが減ってきたが何を食べたいか具体的に決まっていない状況での対話が、目的意識が低い対話となる。前者は候補と提示店名のマッチングであり、後者は提示店名からの連想・判断といえ、目的意識が低い場合でより負荷が高い、と考えられる。

そこで、被験者の目的意識を明確にして実験するため、2種類の実験条件で、対話課題を行う。目的意識[高]では、検索条件にヒットした音声で提示される店名の中から、行きたいお店（事前に指定された店舗候補）を見つけ出し、店舗決定をシステムに伝える。目的意識[低]では、検索条件のみ事前に指定し、検索条件にヒットした音声で提示される店名の中から自由に選択し、行きたいお店をシステムに伝える。それぞれの対話例を図3、図4に示す。

3.2. 実験方法

被験者は大学生の男性7名、女性3名の計10名で、目的意識が高い場合、低い場合のいずれかの実験条件を与え、以下の流れで実験を実施した。

- (1) 反応課題の練習
- (2) 反応課題単体の実施
- (3) 対話課題 (P1), 反応課題の実施
- (4) 対話課題 (P2), 反応課題の実施
- (5) 対話課題 (P3), 反応課題の実施

※ (3) (4) (5) は順不同

各対話後に、NASA-TLX の尺度評価および対話のし易さの主観評価アンケートを行う。そして、実験終了後、NASA-TLX の重み付けのため、評価尺度の一对比較評価を実施する。

3.3. 実験結果

3.3.1. 目的意識が高い場合の負荷量

図5、図6は目的意識が高い場合での、反応遅れ率の平均、およびNASA-TLX 総合値の平均である。反応遅れ率、NASA-TLX 総合値ともに、提示量と比例している事が分かる。つまり、システムが一発話で提示する情報量が増加すると、対話によりユーザに与える負荷量が増えることが実証できた。

また、図7、図8は目的意識が高い場合での被験者ごとの、反応遅れ率およびNASA-TLX 総合値を示しており、反応遅れ率とNASA-TLX 総合値の実験条件間の傾向がおおよそ合致していることが見受けられる。

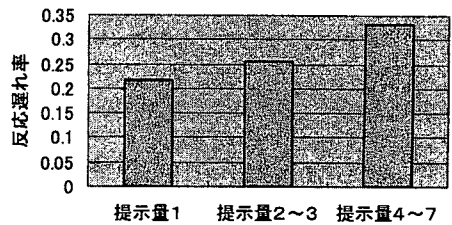


図5 反応遅れ率の平均：目的意識[高]

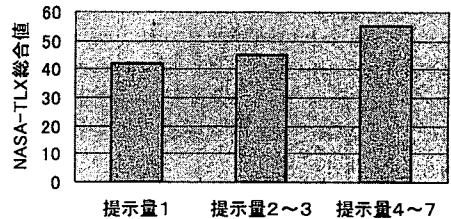


図6 NASA-TLX 総合値の平均：目的意識[高]

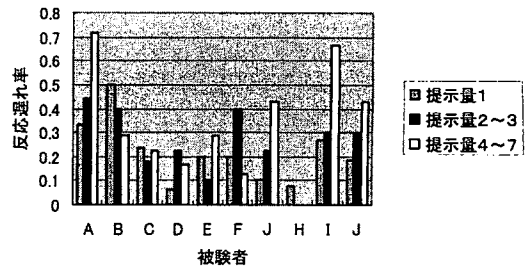


図7 被験者ごとの反応遅れ率：目的意識[高]

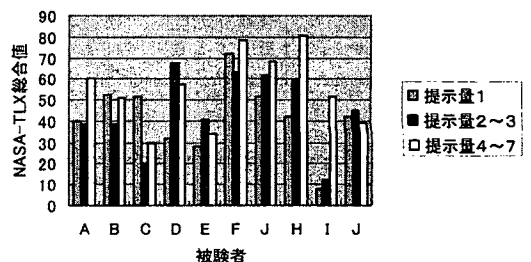


図8 被験者ごとのNASA-TLX 総合値：目的意識[高]

3.3.2. 目的意識が低い場合の負荷量

図9, 図10は目的意識が低い場合での, 反応遅れ率の平均, およびNASA-TLX総合値の平均である. また, 図11, 図12は目的意識が低い場合での被験者ごとの, 反応遅れ率, およびNASA-TLX総合値である.

図9では, 目的意識が高い場合とは異なり, 反応遅れ率と提示量が比例せず, 提示量1, 提示量2~3に比べ, 提示量4~7で反応遅れ率が下がっている. 一方, NASA-TLX総合値に関しては, 目的意識が高い場合と同様, 低い場合でも, NASA-TLX総合値と提示量が比例している.

今回の実験では, 記憶・判断などの思考量が最も多く, 高い負荷がかかると想定していた, 目的意識[低]かつ提示量4~7という条件で, 反応遅れ率がNASA-TLX総合値と異なる傾向が見られた原因を考える. NASA-TLX総合値は他条件に比べても非常に高くなっており, また, 他条件では全員が, 検索候補が全て提示されるまでにお店を決めてタスク達成した一方, 同条件では対話のタスク達成率が7割と低くなっていた. つまり, 一定以上の心的負荷状況では, 対話課題よりも反応課題(=運転)を優先しているとみなす事ができ, これは自然な行動と考えられる. 提示量4~7でのNASA-TLX総合値の平均は, 目的意識[高]では55, 目的意識[低]では67なので, NASA-TLX総合値が70を超えた結果の被験者の反応遅れ率を見ると, 7人中6人で反応遅れ率が抑えられたことが分かる. これは, 運転下において対話が可能な心的負荷の許容量を示しているのではないかと考えられる.

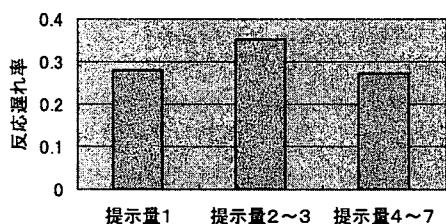


図9 反応遅れ率の平均: 目的意識[低]

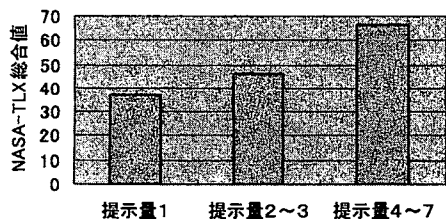


図10 NASA-TLX総合値の平均: 目的意識[低]

3.3.3. 利便性

図13は対話のし易さの主観評価結果である. 負荷量の傾向と異なり, 目的意識が高い場合, 低い場合ともに, 提示量2~3がもっとも話しやすいと評価されている. これは, NASA-TLX評価尺度のフラストレーションの評価(図14)と似た傾向であることが分かる. 一方, NASA-TLXの他の評価尺度では, 提示量と比例する傾向が見受けられた. つまり, 提示量が少なくなると, 店名決定までにかかるターン数が増加する傾向があり, 被験者はフラストレーションを感じ, 対話しにくいと感じてしまうことが分かる. また, 目的意識[低]かつ提示量4~7の条件で対話のし易さの評価が著しく低く, これは, 3.3.2でも述べたような高負荷の対話は, 利便性にも影響を及ぼすことを示している.

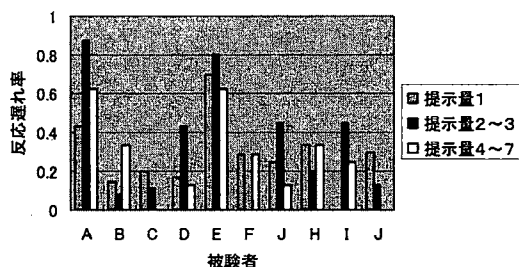


図11 被験者ごとの反応遅れ率: 目的意識[低]

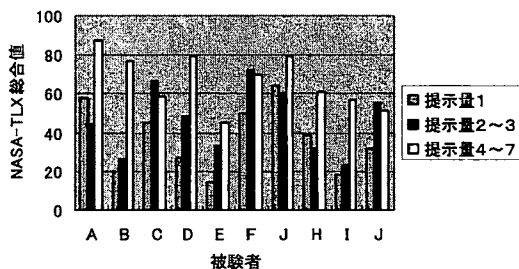


図12 被験者ごとのNASA-TLX総合値: 目的意識[低]

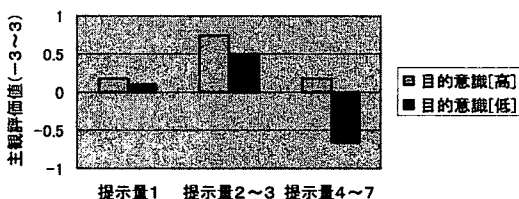


図13 対話のし易さの主観評価値

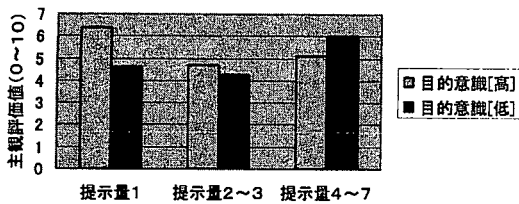


図 1 4 評価尺度：フラストレーション主観評価値

3.4. 考察

車載情報機器を想定した対話戦略では、提示量と負荷量のおおよその相関性より、提示量を増加させると対話が与えるユーザへの影響が高くなり、運転に悪い影響が生じるため、平常時以上に提示量を抑える必要があると考えられる。しかし、提示量を抑えすぎると、ユーザが必要とする情報にたどり着くターン数が増加する可能性があるため、利便性が下がり、ユーザが音声対話システムを使用しなくなる可能性がある。特に目的意識が高い場合、提示量が1ではユーザに強くフラストレーションを与えてしまう。また、NASA-TLXの評価値が運転時における、対話が可能な心的負荷の許容量を示している可能性が明らかとなり、今後、心的負荷を抑えた対話戦略を検討する上での尺度の一つとして、さらに検討を行う。

以上の事から、車載情報機器の対話戦略例として、ブレーキ、アクセル、ウインカーなどの情報から運転状況を推定し、運転にかかる負担が低い定常運転時では、提示量を2~3個に、運転状況に応じて提示量を1、更には対話を一旦打ち切るなどの対応をすることにより、利便性と安全性の両方を実現できると考えられる。更に、車載情報機器は同一ユーザが使い続ける傾向にあり、ユーザモデルを使用するのに適した環境であるため、ユーザモデルを用いて、ユーザの目的意識推定による提示量の変更や、嗜好に合わせた提示順序変更による要求情報の素早い提示などができれば、利便性が下がる問題点を解決できると考えられる。

4. おわりに

心的負荷状況における、システムとの対話がユーザに与える影響を考慮した音声対話戦略の提案と、提案を実装した音声対話システムの実現を目指し、一発話で提示する情報量や検索対話での目的意識の差が、ユーザに与える影響を調べるため、二重課題法とNASA-TLXを用い、負荷測定実験を行った。その結果、システムの一発話での提示量と負荷量が比例していることが実証された。また、対話のし易さの主観評価値や、NASA-TLXの評価尺度分析から、提示量を抑えす

ぎると、利便性が下がることが確認できた。この結果をもとに、車載情報機器の対話戦略に関する考察を行った。

今後は、考察した車載情報機器を想定した音声対話戦略をシステムに実装し、より実運転に近い環境で評価を進めていく予定である。そして、利便性・安全性を備えた対話システムの実現を進めていく。

謝辞 本研究は、富士重工業株式会社との共同研究により実施した。

文 献

- [1] 小田達也, 桐山伸也, 北澤茂良, “食事シチュエーションにおける気の利いた状況理解と情報提示による快走支援”, 人工知能学会 第 20 回全国大会, June 2006.
- [2] 宗近純一, 松坂要佐, 小林哲則, “車運転時における心的作業負荷”, 第 2 回情報科学技術フォーラム FIT2003 情報技術レターズ, vol. 2, pp.105-106, Sep 2003.
- [3] 脇田敏裕, 寺島立太, 小島真一, 清水司, 本郷武朗, “運転中情報機器操作性の評価法”, 情報処理学会論文誌, vol.42, No7, pp.1762-1769, July 2001.
- [4] 清水司, 小島真一, 脇田敏裕, 本郷武朗, “運転中における音声対話システムの評価”, 情報処理学会研究会報告, 2000-SLP-32-17, pp.87-92, July 2000.
- [5] 西本卓也, 高山元希, 櫻井晴章, 荒木雅弘, “音声インタフェースのための対話負荷測定法”, 電子情報通信学会論文誌, vol.J87-D-II, No2, pp.513-520, Feb 2004.
- [6] 池田拓郎, 北原俊夫, 岩田収, 福岡俊之, “車室内における音声インタフェースのユーザ満足度向上に向けた実験的考察”, ヒューマンインタフェースシンポジウム 2006, 論文集, pp.613-616.
- [7] 内山祐司, 小島真一, 本郷武朗, 寺島立太, 脇田敏裕 “運転状況適応型音声情報提示システム”, ヒューマンインタフェースシンポジウム 2000, 論文集, pp.375-378.
- [8] 萩原亨, 徳永ロベルト, “メンタルワークロード評価法に基づく携帯電話利用の影響に関する研究”, 国際交通安全学会誌, vol.30, No3, pp.66-73 Sep 2005.
- [9] 篠塚一光, 中村隆宏, 竜田成示, 井場陽一, “多視点ディスプレイを用いた車載情報機器利用により生じる注意転動の評価”, ヒューマンインタフェースシンポジウム 2006, pp.1079-1082.
- [10] 工藤貴弘, 小澤順, 吉岡元貴, “アバタとエージェントを利用した仮想対話インタフェースによる Soft Interaction”, 人工知能学会論文誌, 19 巻 4 号 I, pp.351-359, 2004.
- [11] 清信将範, 川端豪, “ドライビングシミュレータの走行精度による車載情報機器の負荷測定”, 信学技報 SP2006-13, pp.13-18, July 2006.
- [12] 中園薫, 竹田祐樹, 寺内美奈, 長嶋祐二, “遅延環境下における映像対話のメンタルワークロードの評価”, 信学技報, WIT2006-49, pp.95-100, Oct 2006.
- [13] 芳賀繁 “メンタルワークロードの理論と測定”, 日本出版サービス, 2001.